

FIG. 1. Parallelization Method

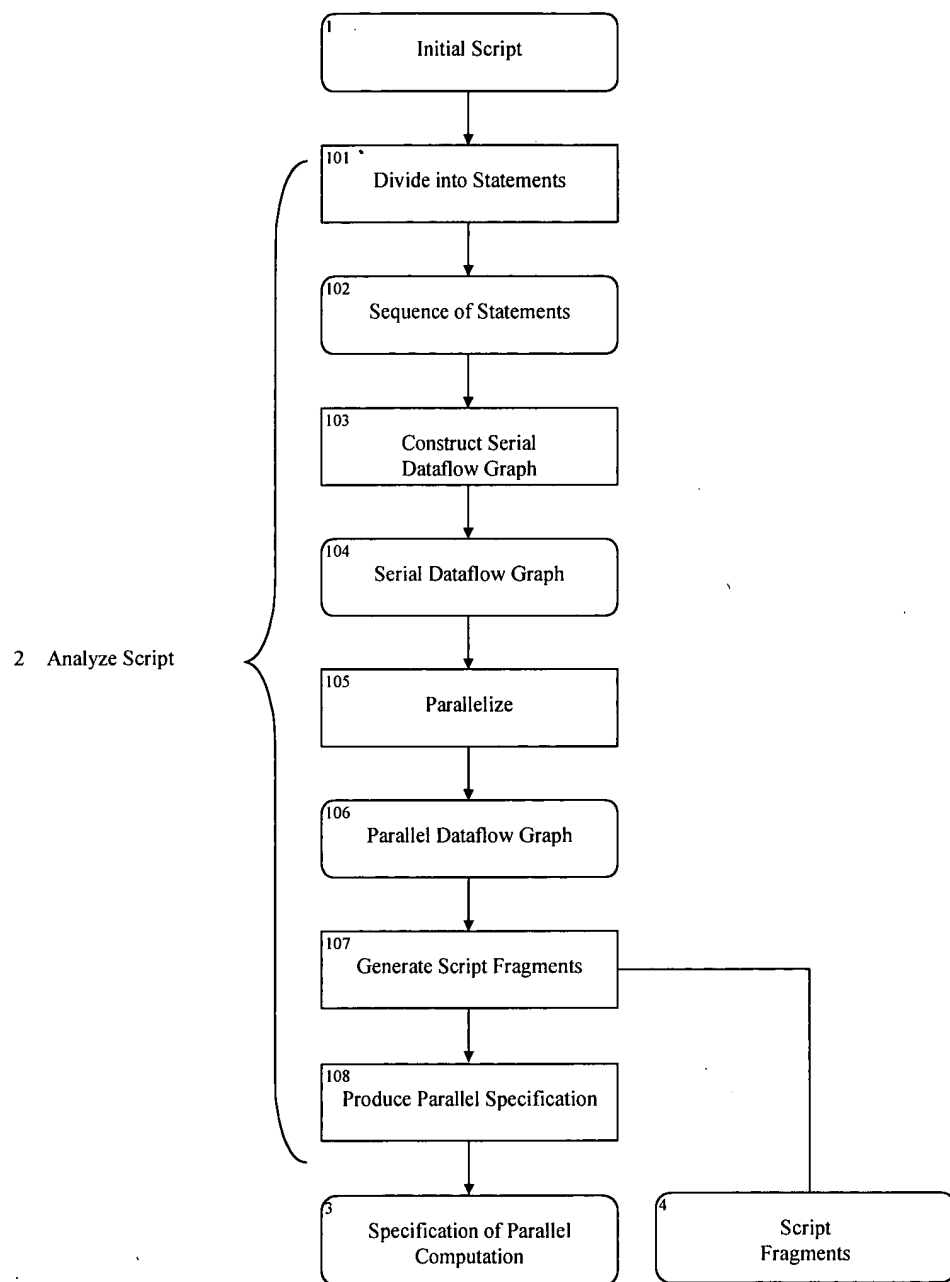


FIG. 2. Method for Analyzing Scripts

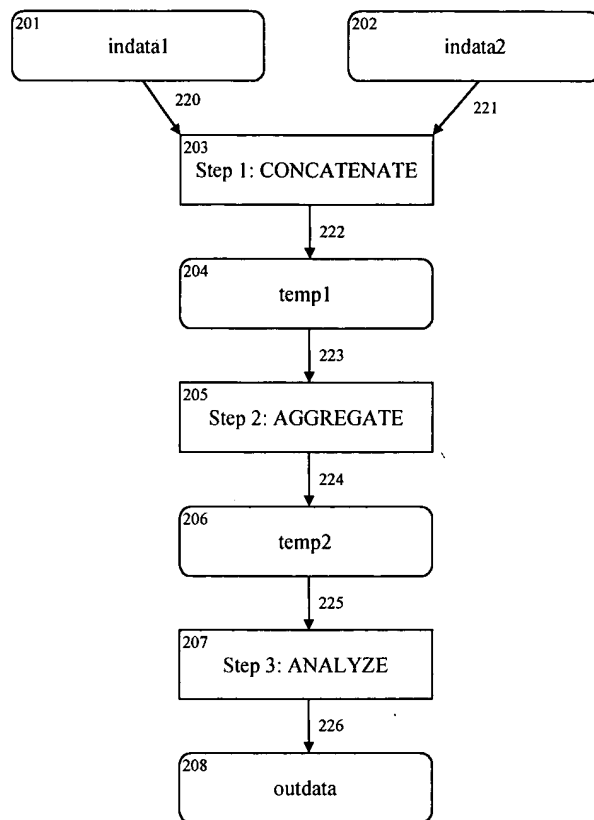


FIG. 3. Serial dataflow graph

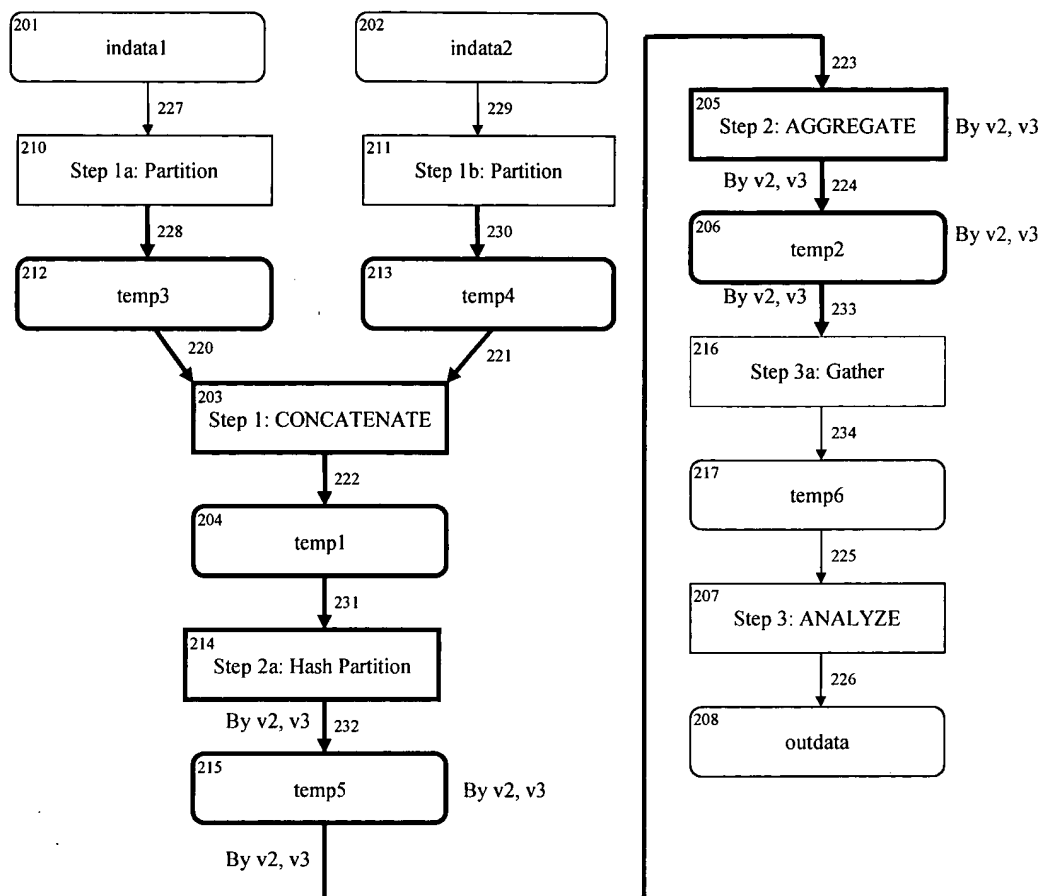


FIG. 4. Parallel Dataflow Graph

1

```
INPUT indatal input1.dat
INPUT indata2 input2.dat
CONCATENATE indatal indata2
AGGREGATE o1 = MIN v1, o2 = MAX v1, o3 = SUM v1 BY v2, v3
ANALYZE
OUTPUT outdata output.dat
```

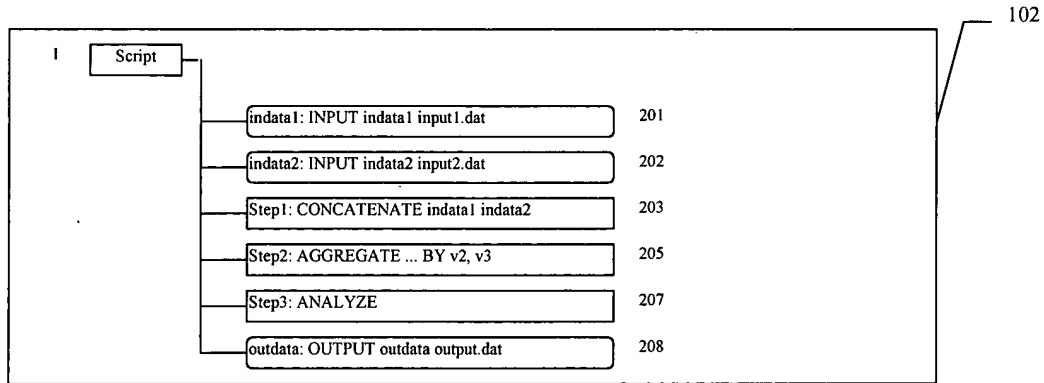


FIG. 5. Original Script and division into processing steps

	Name	Class	File	Explicit?	
201	indata1	INPUT	input1.dat	Yes	601
202	indata2	INPUT	input2.dat	Yes	
208	outdata	OUTPUT	output.dat	Yes	
204	temp1	TEMP		No	
206	temp2	TEMP		No	
Data Set Table					

	Name	Operation	Parameteres (excepting data sets)	
203	Step1	CONCATENATE		602
205	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	
207	Step3	ANALYZE		
Processing Step Table				

	Step	Data Set	Direction	Role	
220	Step1	indata1	input	in	603
221	Step1	indata2	input	in	
222	Step1	temp1	output	out	
223	Step2	temp1	input	in	
224	Step2	temp2	output	out	
225	Step3	temp2	input	in	
226	Step3	outdata	output	out	
Data Set Access Table					

FIG. 6. Serial Dataset, Processing Step, and Dataset Access Tables

201 202 208 204 206 203 205 207 220 221 222 223 224 225 226

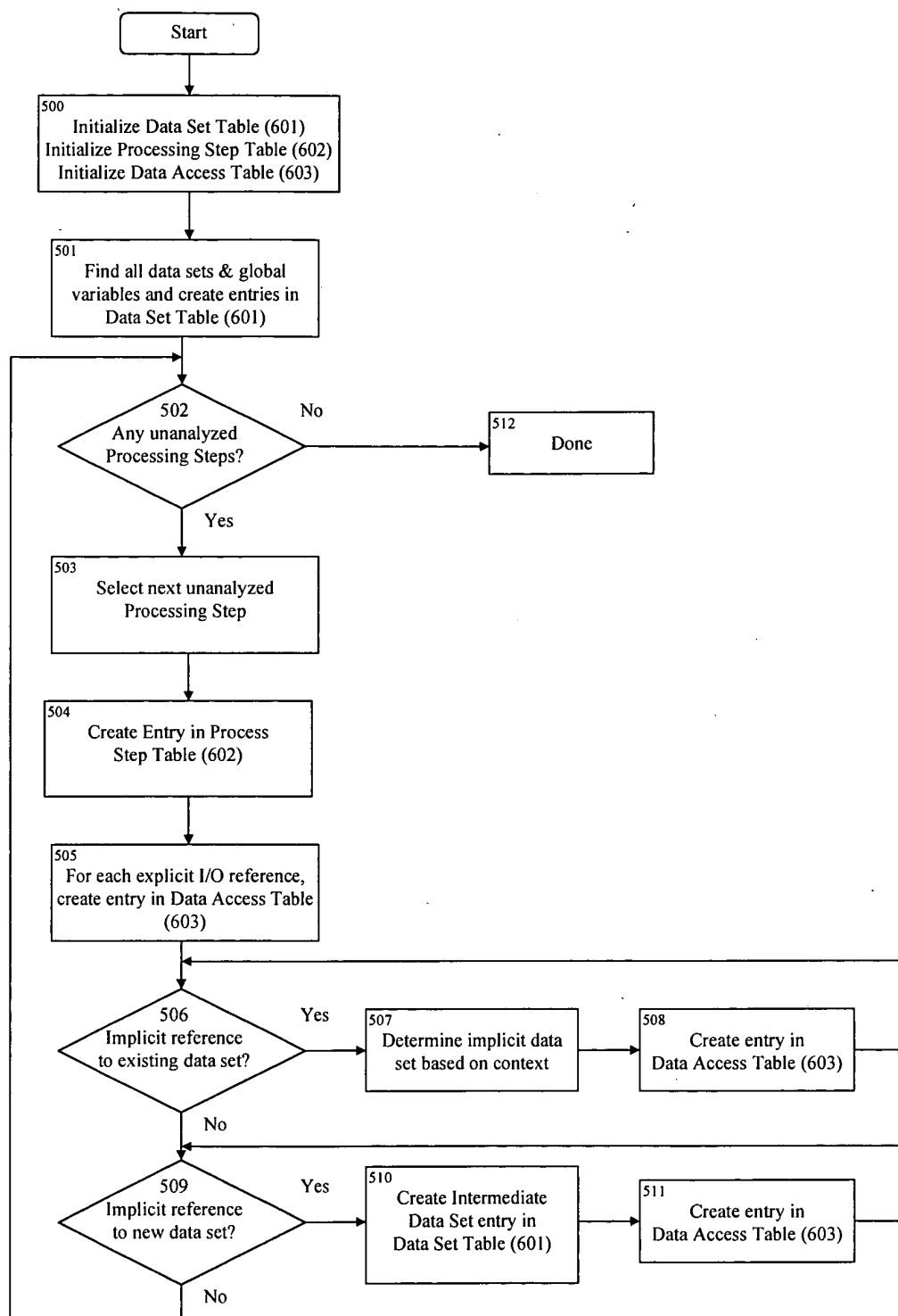


FIG. 7. Method for converting sequences of steps to serial dataflow graph

901	902
903 CONCAT or COPY processing step	Parallelize by simple partitioning
904 Processing step contains BY clause	Parallelize by key-based partitioning
905 INPUT, OUTPUT, or TEMP data set accessing partitioned file	Parallelize by expansion of the partitioned file.
906 AGGREGATE operation	Parallelize by splitting into global and local stages.
907 IN_DB data set accessing partitioned database table.	Parallelize by adding an external "db unload" process.
908 OUT_DB data set accessing partitioned database table	Parallelize by adding an external "db load" process.

FIG. 8. Parallelization repertoire

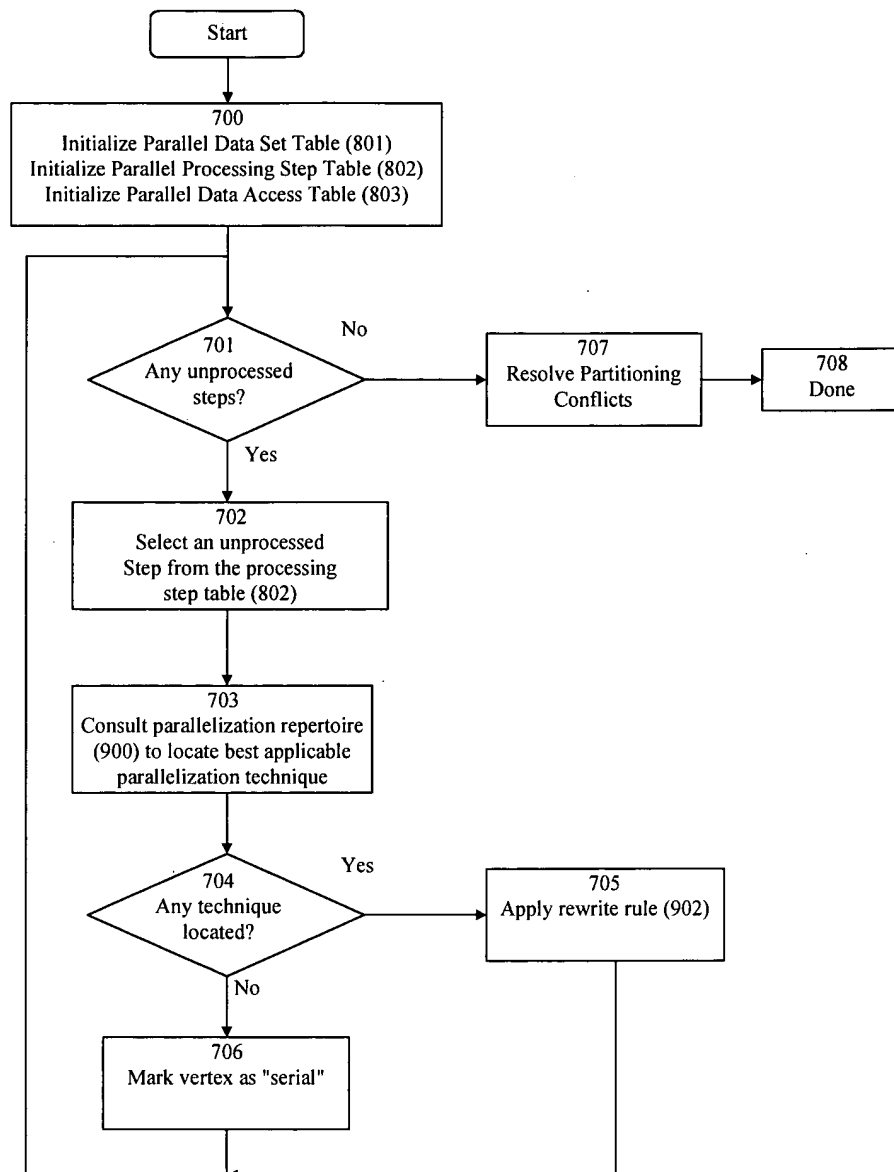


FIG. 9. Method for Parallelizing serial dataflow graph

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	801
206	temp1	TEMP		No		
	temp2	TEMP		No		
	Parallel Data Set Table (initial)					

203	Name	Operation	Parameters (excepting data sets)	Partitioning	
205	Step1	CONCATENATE			
207	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3		802
	Step3	ANALYZE			
	Parallel Processing Step Table (initial)				

220	Step	Data Set	Direction	Role	Partitioning	
221	Step1	indata1	input	in		
222	Step1	indata2	input	in		
223	Step1	temp1	output	out		803
224	Step2	temp1	input	in		
225	Step2	temp2	output	out		
226	Step3	temp2	input	in		
	Step3	outdata	output	out		
	Parallel Data Set Access Table (initial)					

FIG. 10 Initial Parallel Dataset, Processing Step, and Dataset Access Tables

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	801
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	
206	temp1	TEMP		No		
	temp2	TEMP		No		
	Parallel Data Set Table					

203	Name	Operation	Parameters (excepting data sets)	Partitioning	
205	Step1	CONCATENATE		Simple	802
	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3	
207	Step3	ANALYZE		Serial	
	Parallel Processing Step Table				

220	Step	Data Set	Direction	Role	Partitioning	
221	Step1	indata1	input	in	Simple	803
222	Step1	indata2	input	in	Simple	
223	Step1	temp1	output	out	Simple	
224	Step2	temp1	input	in	By v2, v3	
225	Step2	temp2	output	out	By v2, v3	
226	Step3	temp2	input	in	Serial	
	Step3	outdata	output	out	Serial	
	Parallel Data Set Access Table					

FIG. 11. Parallel Dataset, Processing Step, and Dataset Access Tables, after parallelizing processing steps.

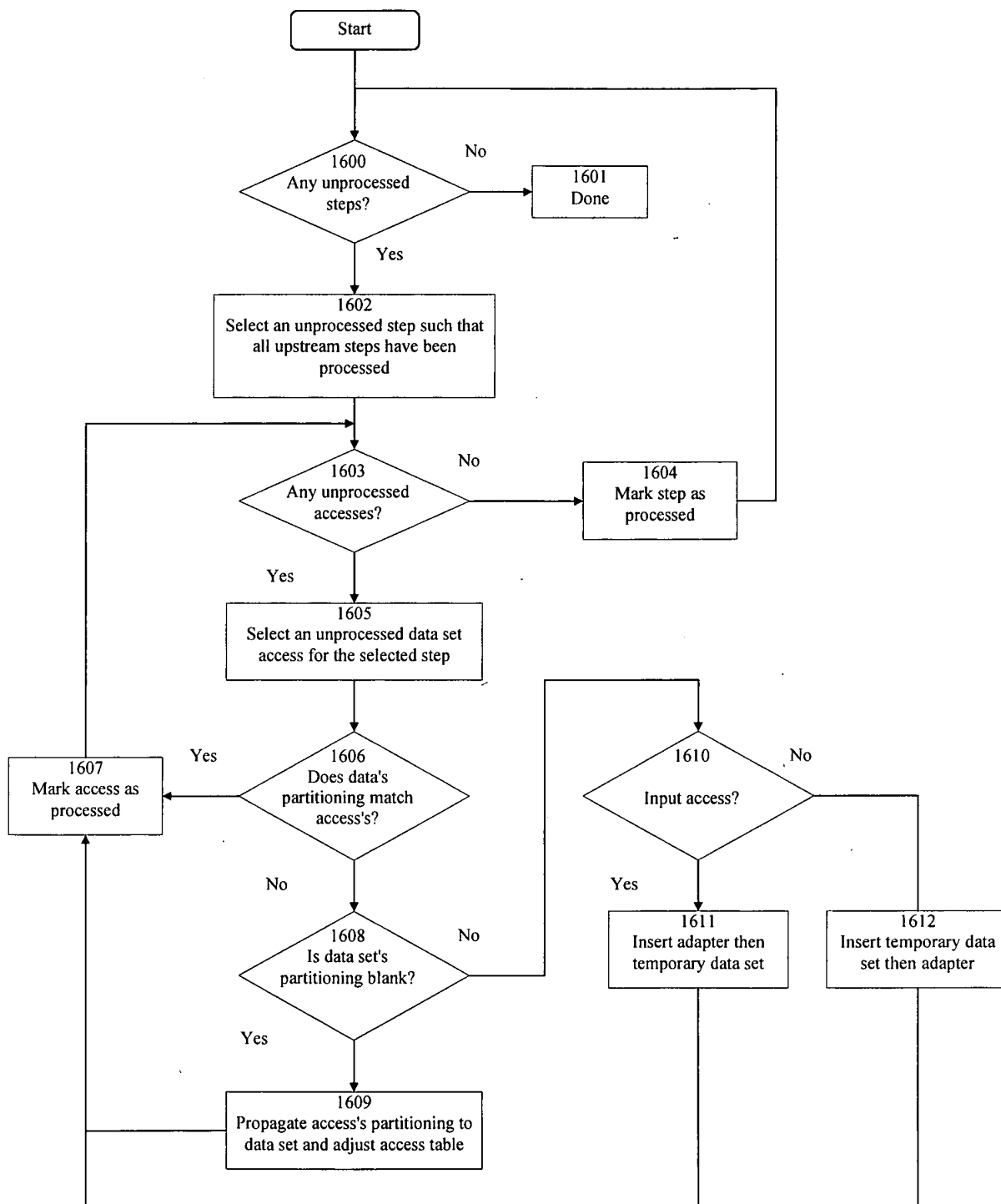


FIG. 12. Method for resolving partitioning conflicts.

201	Name	Class	File	Explicit?	Partitioning	801
202	indata1	INPUT	input1.dat	Yes	Serial	
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	
206	temp1	TEMP		No	Simple	
212	temp2	TEMP		No	Simple	
213	temp3	TEMP		No	Simple	
Parallel Data Set Table						

210	Name	Operation	Parameters (excepting data sets)	Partitioning	802
211	Step1a	Simple-Partition		Serial	
203	Step1b	Simple-Partition		Serial	
205	Step1	CONCATENATE		Simple	
207	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3	
	Step3	ANALYZE		Serial	
Parallel Processing Step Table					

227	Step	Data Set	Direction	Role	Partitioning	803
228	Step1a	indata1	input	in	Serial	
229	Step1a	temp3	output	out	Simple	
230	Step1b	indata2	input	in	Serial	
220	Step1b	temp4	output	out	Simple	
221	Step1	temp3	input	in	Simple	
222	Step1	temp4	input	in	Simple	
223	Step1	temp1	output	out	Simple	
224	Step2	temp1	input	in	By v2, v3	
225	Step2	temp2	output	out	By v2, v3	
226	Step3	temp2	input	in	Serial	
	Step3	outdata	output	out	Serial	
Parallel Data Set Access Table						

FIG. 13. Parallel Dataset, Processing Step, and Dataset Access Tables, after resolving Step 1

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	801
206	temp1	TEMP		No	Simple	
212	temp2	TEMP		No	By v2,v3	
213	temp3	TEMP		No	Simple	
215	temp4	TEMP		No	Simple	
	temp5	TEMP		No	By v2,v3	
	Parallel Data Set Table					

210	Name	Operation	Parameters (excepting data sets)	Partitioning	
211	Step1a	Simple-Partition		Serial	
203	Step1b	Simple-Partition		Serial	
214	Step1	CONCATENATE		Simple	
205	Step2a	Hash-Partition	By v2, v3	Simple	
207	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3	
	Step3	ANALYZE		Serial	
	Parallel Processing Step Table				802

227	Step	Data Set	Direction	Role	Partitioning	
228	Step1a	indata1	input	in	Serial	
229	Step1a	temp3	output	out	Simple	
230	Step1b	indata2	input	in	Serial	
220	Step1b	temp4	output	out	Simple	
221	Step1	temp3	input	in	Simple	803
222	Step1	temp4	input	in	Simple	
231	Step1	temp1	output	out	Simple	
232	Step2a	temp1	input	in	Simple	
223	Step2a	temp5	output	out	By v2,v3	
224	Step2	temp5	input	in	By v2,v3	
225	Step2	temp2	output	out	By v2,v3	
226	Step3	temp2	input	in	Serial	
	Step3	outdata	output	out	Serial	
	Parallel Data Set Access Table					

FIG. 14. Parallel Dataset, Processing Step, and Dataset Access Tables, after resolving Step 2

	Name	Class	File	Explicit?	Partitioning
201	indata1	INPUT	input1.dat	Yes	Serial
202	indata2	INPUT	input2.dat	Yes	Serial
208	outdata	OUTPUT	output.dat	Yes	Serial
204	temp1	TEMP		No	Simple
206	temp2	TEMP		No	By v2,v3
212	temp3	TEMP		No	Simple
213	temp4	TEMP		No	Simple
215	temp5	TEMP		No	By v2,v3
217	temp6	TEMP		No	Serial
Parallel Data Set Table					

	Name	Operation	Parameters (excepting data sets)	Partitioning
210	Step1a	Simple-Partition		Serial
211	Step1b	Simple-Partition		Serial
203	Step1	CONCATENATE		Simple
214	Step2a	Hash-Partition	By v2, v3	Simple
205	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3
216	Step3a	Gather		Serial
207	Step3	ANALYZE		Serial
Parallel Processing Step Table				

	Step	Data Set	Direction	Role	Partitioning
227	Step1a	indata1	input	in	Serial
228	Step1a	temp3	output	out	Simple
229	Step1b	indata2	input	in	Serial
230	Step1b	temp4	output	out	Simple
220	Step1	temp3	input	in	Simple
221	Step1	temp4	input	in	Simple
222	Step1	temp1	output	out	Simple
231	Step2a	temp1	input	in	Simple
232	Step2a	temp5	output	out	By v2,v3
223	Step2	temp5	input	in	By v2,v3
224	Step2	temp2	output	out	By v2,v3
233	Step3a	temp2	input	in	By v2,v3
234	Step3a	temp6	output	out	Serial
225	Step3	temp6	input	in	Serial
226	Step3	outdata	output	out	Serial
Parallel Data Set Access Table					

FIG. 15. Parallel Dataset, Processing Step, and Dataset Access Tables, after resolving Step 3

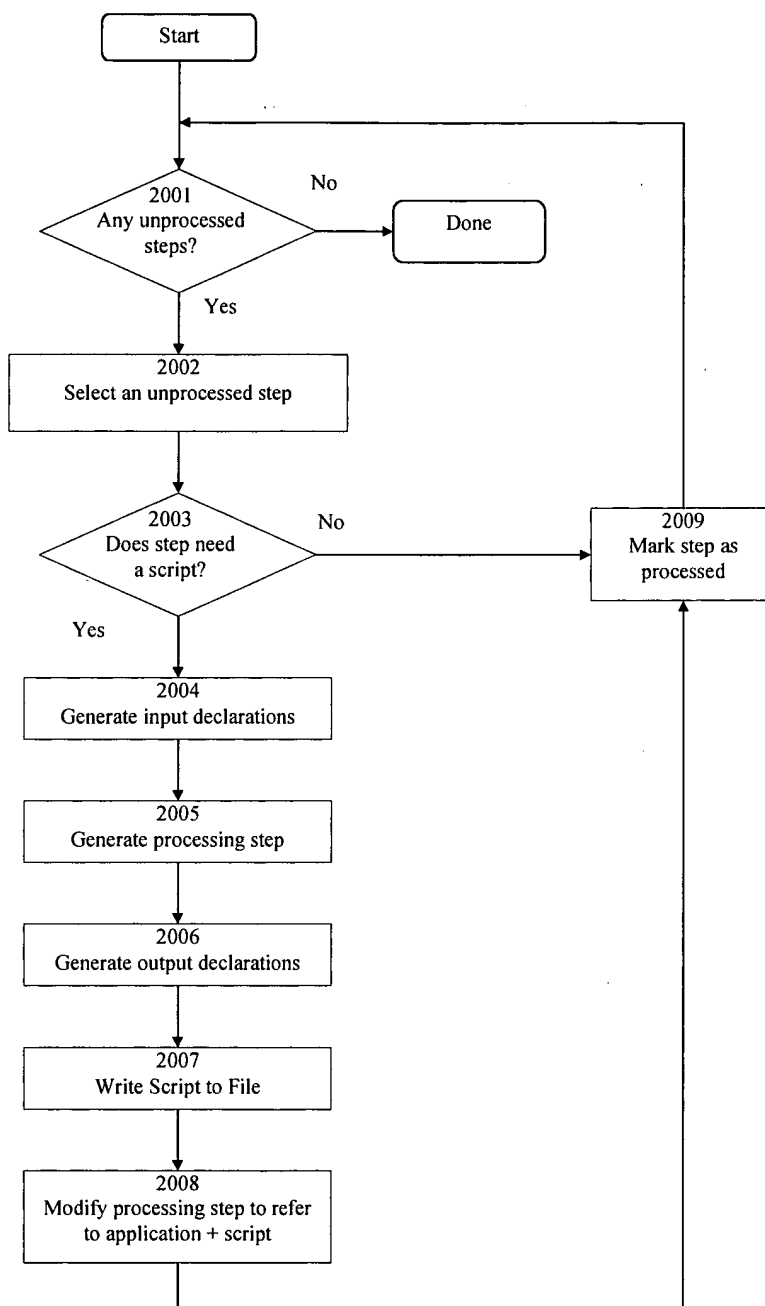


FIG. 16. Method for generating script fragments

	Name	Class	File	Explicit?	Partitioning	
201	indata1	INPUT	input1.dat	Yes	Serial	801
202	indata2	INPUT	input2.dat	Yes	Serial	
208	outdata	OUTPUT	output.dat	Yes	Serial	
204	temp1	TEMP		No	Simple	
206	temp2	TEMP		No	By v2,v3	
212	temp3	TEMP		No	Simple	
213	temp4	TEMP		No	Simple	
215	temp5	TEMP		No	By v2,v3	
217	temp6	TEMP		No	Serial	
Parallel Data Set Table						

	Name	Operation	Parameters (excepting data sets)	Partitioning	
210	Step1a	Simple-Partition		Serial	802
211	Step1b	Simple-Partition		Serial	
203	Step1	Run analyze	script1	Simple	
214	Step2a	Hash-Partition	By v2, v3	Simple	
205	Step2	Run analyze	script2	By v2,v3	
216	Step3a	Gather		Serial	
207	Step3	Run analyze	script3	Serial	
Parallel Processing Step Table					

	Step	Data Set	Direction	Role	Partitioning	
227	Step1a	indata1	input	in	Serial	803
228	Step1a	temp3	output	out	Simple	
229	Step1b	indata2	input	in	Serial	
230	Step1b	temp4	output	out	Simple	
220	Step1	temp3	input	in	Simple	
221	Step1	temp4	input	in	Simple	
222	Step1	temp1	output	out	Simple	
231	Step2a	temp1	input	in	Simple	
232	Step2a	temp5	output	out	By v2,v3	
223	Step2	temp5	input	in	By v2,v3	
224	Step2	temp2	output	out	By v2,v3	
233	Step3a	temp2	input	in	By v2,v3	
234	Step3a	temp6	output	out	Serial	
225	Step3	temp6	input	in	Serial	
226	Step3	outdata	output	out	Serial	
Parallel Data Set Access Table						

FIG. 18. Parallel Dataset, Processing Step, and Dataset Access Tables, after generation of script fragments

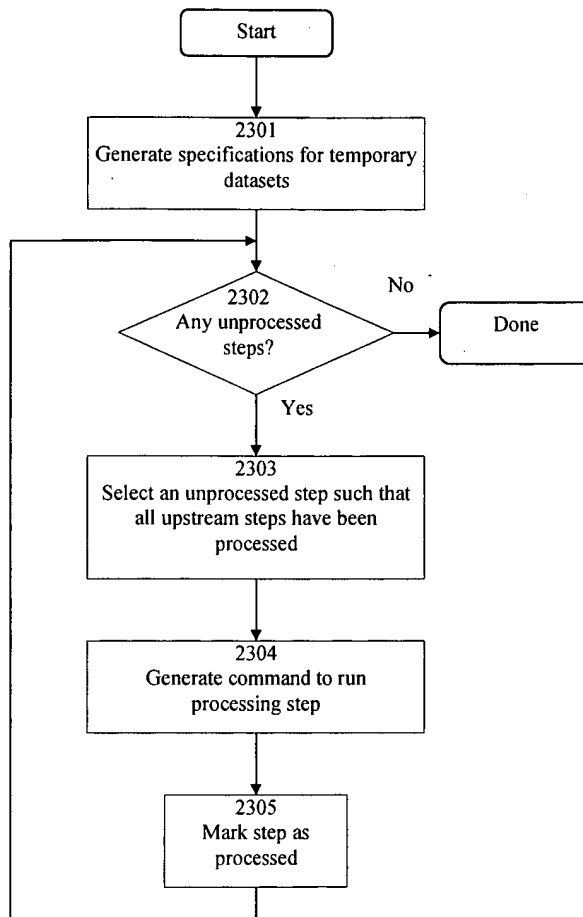


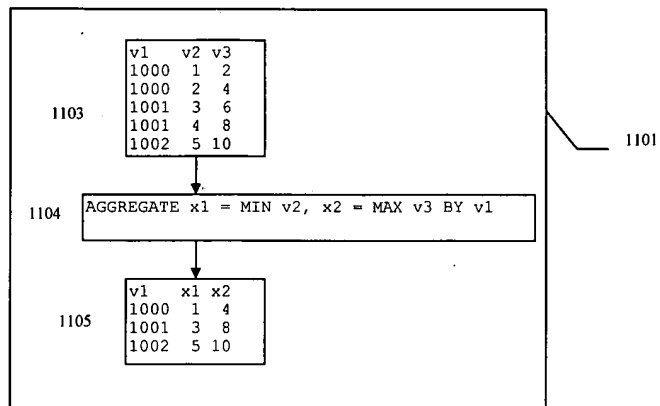
FIG. 19. Method for generating parallel specification

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	801
206	temp1	TEMP	temp1.1;temp1.2	No	Simple	
212	temp2	TEMP	temp2.1;temp2.2	No	By v2,v3	
213	temp3	TEMP	temp3.1;temp3.2	No	Simple	
215	temp4	TEMP	temp4.1;temp4.2	No	Simple	
217	temp5	TEMP	temp5.1;temp5.2	No	By v2,v3	
	temp6	TEMP	temp6	No	Serial	
	Parallel Data Set Table					

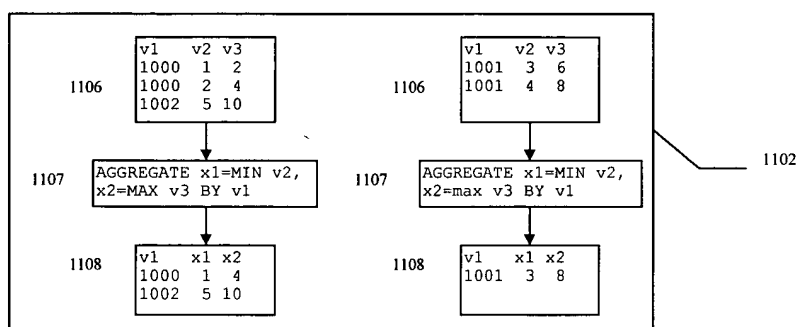
210	Name	Operation	Parameters (excepting data sets)	Partitioning	
211	Step1a	Simple-Partition		Serial	
203	Step1b	Simple-Partition		Serial	
214	Step1	run analyze	script1	Simple	802
205	Step2a	Hash-Partition	By v2, v3	Simple	
216	Step2	run analyze	script2		
207	Step3a	Gather		Serial	
	Step3	run analyze	script3	Serial	
	Parallel Processing Step Table				

227	Step	Data Set	Direction	Role	Partitioning	
228	Step1a	indata1	input	in	Serial	
229	Step1a	temp3	output	out	Simple	
230	Step1b	indata2	input	in	Serial	
220	Step1b	temp4	output	out	Simple	803
221	Step1	temp3	input	in	Simple	
222	Step1	temp4	input	in	Simple	
231	Step1	temp1	output	out	Simple	
232	Step2a	temp1	input	in	Simple	
223	Step2a	temp5	output	out	By v2,v3	
224	Step2	temp5	input	in	By v2,v3	
233	Step2	temp2	output	out	By v2,v3	
234	Step3a	temp2	input	in	By v2,v3	
225	Step3a	temp6	output	out	Serial	
226	Step3	temp6	input	in	Serial	
	Step3	outdata	output	out	Serial	
	Parallel Data Set Access Table					

FIG. 20. Parallel Dataset, Processing Step, and Dataset Access Tables, after defining temporary datasets



Serial Version



Parallel Version

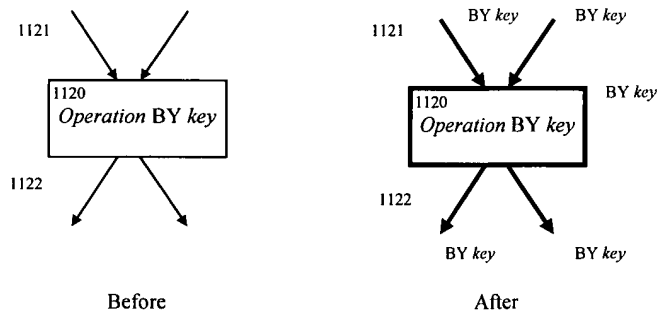
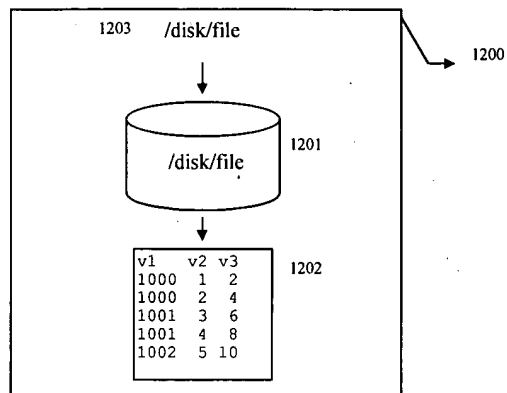
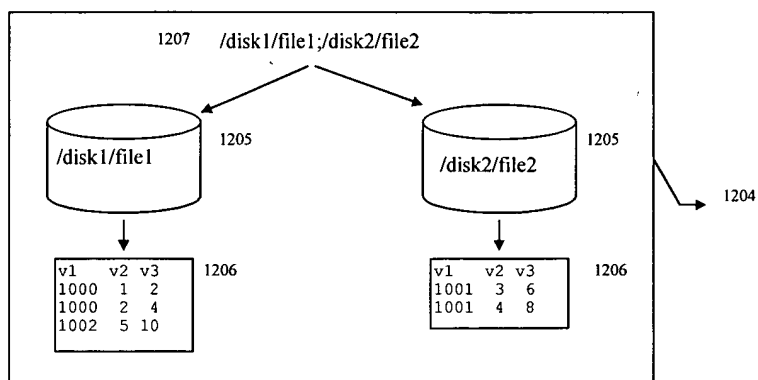


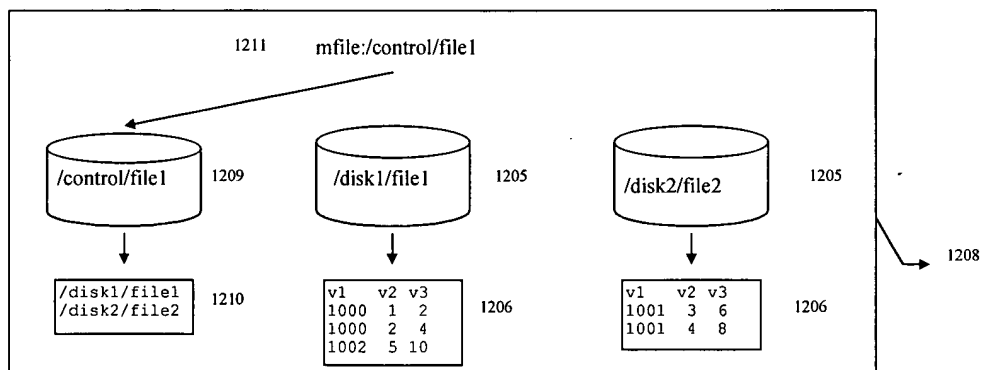
FIG. 23. Key-based Partitioning



Serial Dataset



Partitioned Dataset



Partitioned Dataset with Parallel Virtual File System

FIG. 24. Access to Partitioned Datasets

66270-0730220

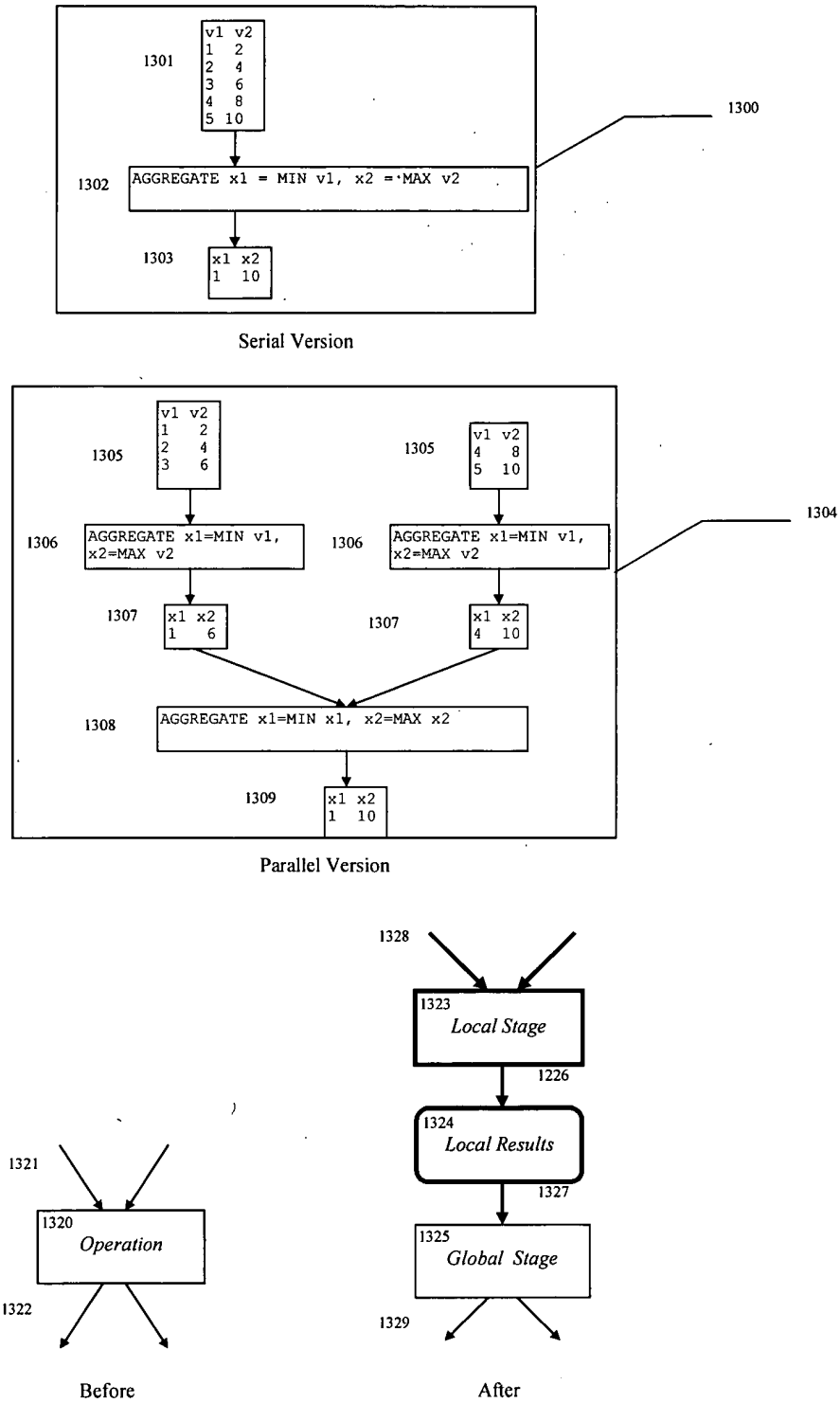


FIG. 25. Local-Global Division Parallelization

60270-5486250

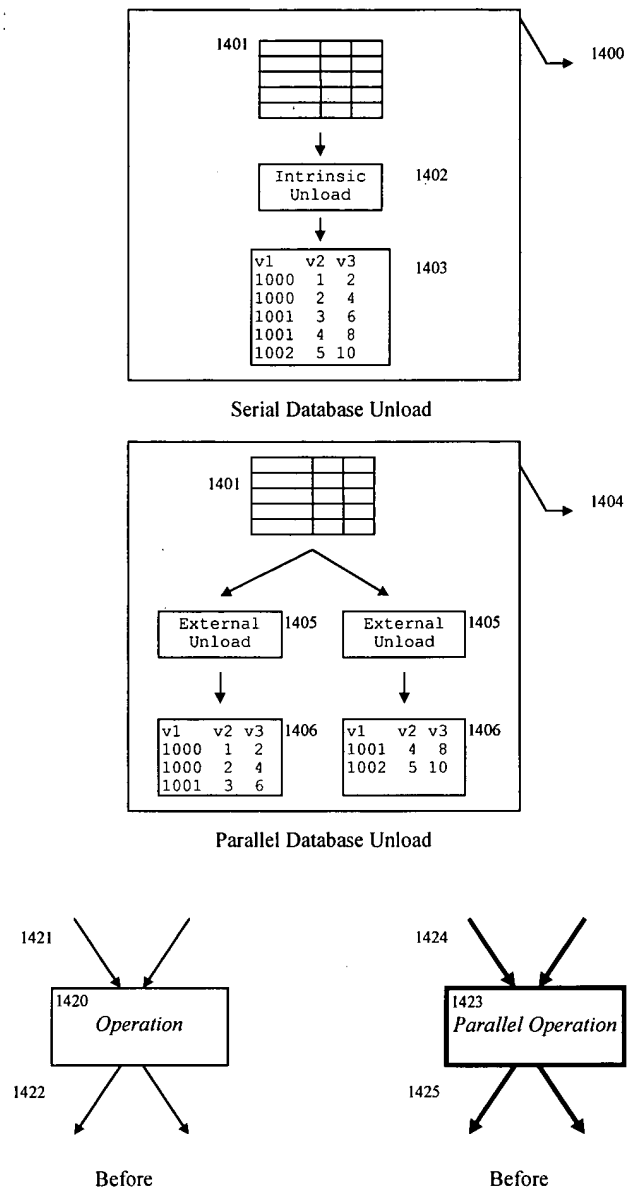


FIG. 26. External Parallelism

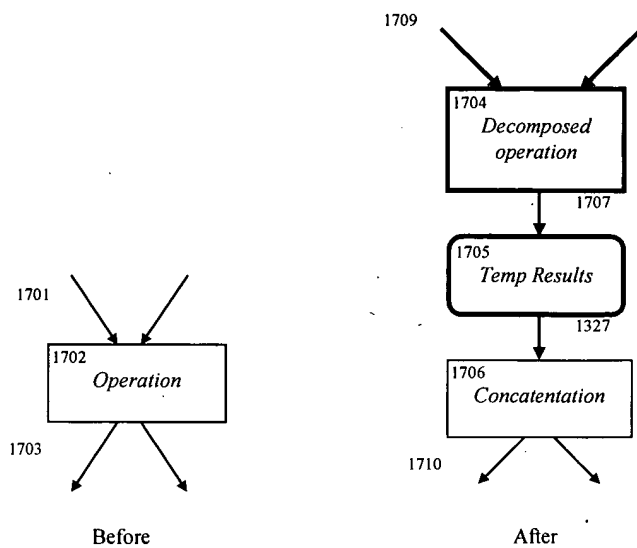


FIG. 27. Statement Decomposition

66770-0786260

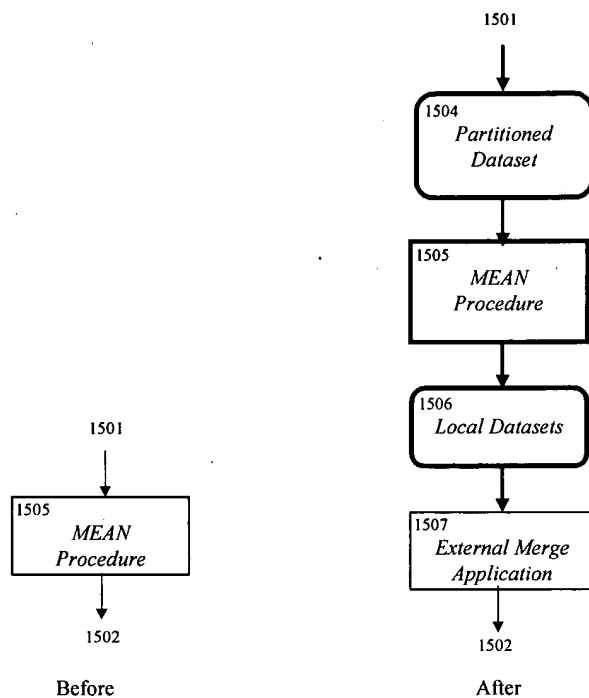


FIG. 28. Local-Global Division for SAS MEANS process

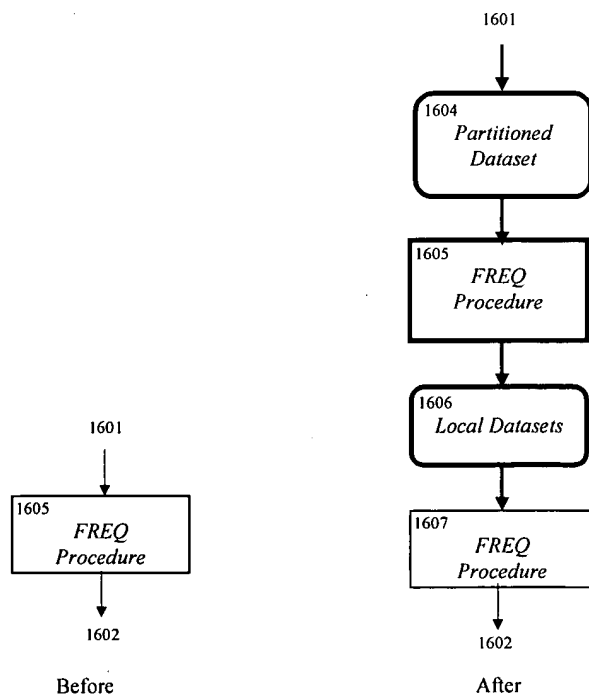


FIG. 29. Local-Global Division for SAS FREQ process

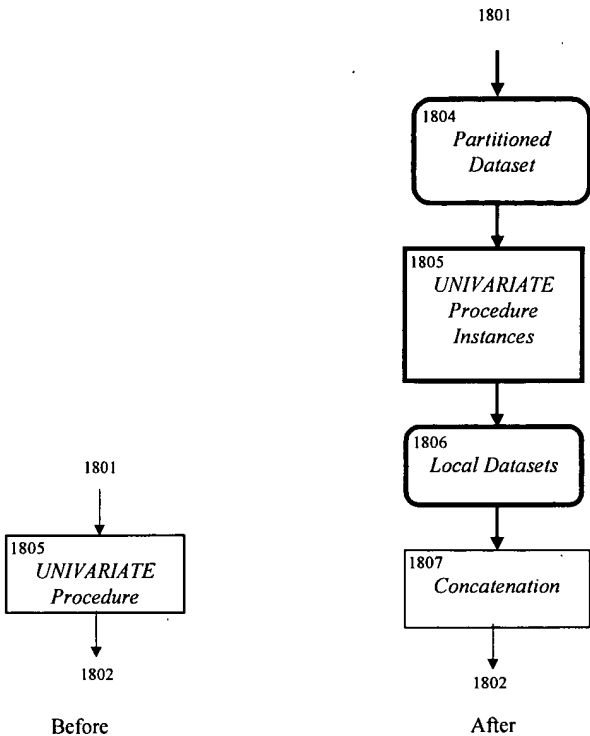
[illegible]

FIG. 30. Statement Decomposition for SAS UNIVARIATE process